Crystal structure of chloroplast cytochrome *f* reveals a novel cytochrome fold and unexpected heme ligation

Sergio E Martinez, Deru Huang, Andrzej Szczepaniak†, William A Cramer* and Janet L Smith*

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA

Background: Cytochrome *f* is the high potential electron acceptor of the chloroplast cytochrome b_6f complex, and is the electron donor to plastocyanin. The 285-residue cytochrome *f* subunit is anchored in the thylakoid membrane of the chloroplast by a single membrane-spanning segment near the carboxyl terminus. A soluble redox-active 252-residue lumen-side polypeptide with native spectroscopic and redox properties, missing the membrane anchor and carboxyl terminus, was purified from turnip chloroplasts for structural studies.

Results: The crystal structure of cytochrome *f*, determined to 2.3 Å resolution, has several unexpected features. The 252-residue polypeptide is organized into one large and one small domain. The larger heme-binding domain is strikingly different from known structures of other *c*-type cytochromes and has the same fold as the type III domain of the animal protein, fibronectin. Cytochrome *f* binds heme with an unprecedented axial heme iron ligand: the amino terminus of the polypeptide.

Conclusion: The first atomic structure of a subunit of either the cytochrome b_6f complex or of the related cytochrome bc_1 complex has been obtained. The structure of cytochrome *f* allows prediction of the approximate docking site of plastocyanin and should allow systematic studies of the mechanism of intra- and inter-protein electron transfer between the cytochrome heme and plastocyanin copper, which are approximately isopotential. The unprecedented axial heme iron ligand also provides information on the sequence of events (i.e. cleavage of signal peptide and ligation of heme) associated with translocation of the cytochrome across the membrane and its subsequent folding.

Structure 15 February 1994, **2**:95–105

Key words: anomalous scattering, cytochrome b_6f, heme iron ligand, membrane protein, protein translocation

Introduction

The cytochrome b_6f complex is one of three integral membrane protein complexes involved in electron transport in oxygenic photosynthesis. The b_6f complex occupies an electrochemically central position in the non-cyclic electron transport chain. The complex accepts electrons from the oxygen-producing photosystem II reaction center, donates electrons to the nicotinamide adenine dinucleotide phosphate ($NADP^+$)-reducing photosystem I reaction center, and contributes to the electrochemical proton gradient that drives ATP synthesis [1]. The redox partners of cytochrome *f* are its donor, the Rieske iron–sulfur protein of the b_6f complex, and its acceptor, the soluble copper protein plastocyanin, which carries electrons to the photosystem I reaction center. The b_6f complex resembles the cytochrome bc_1 complex of the mitochondrial respiratory chain and of purple photosynthetic bacteria in its general energy-transducing function, and in a number of properties of the complex and of its individual polypeptides [2].

Cytochrome *f* is the largest (285 residues, 31 298 Da in turnip chloroplasts) of the four polypeptides with $M_r > 15\,000$ that constitute the chloroplast cytochrome b_6f complex [3], the others being the heme-containing cytochrome b_6 (M_r 23 000), the [2Fe–2S] Rieske protein (M_r 20 000), and the cofactor-free subunit IV (M_r 18 000) [4]. The intact complex contains one copy of each of the four larger polypeptides and is an integral protein complex of the chloroplast thylakoid membrane. Cytochrome *f* contains the characteristic fingerprint sequence Cys-X-Y-Cys-His of *c*-type cytochromes, which is responsible for covalent attachment of the heme group.

Cytochrome *f*, like cytochrome c_1, is anchored to the membrane by one hydrophobic transmembrane segment (residues 251–270) near the carboxyl terminus so that most of the protein (residues 1–250) is located on the lumen side of the thylakoid membrane (Fig. 1a) [3,5,6]. Cytochrome *f* from turnip, a member of the *Cruciferae* family, was chosen for structural work because the purified protein from this source was reported to be soluble and monomeric [7]. Carboxy-terminal sequencing [8] and electrospray mass spectrometry (unpublished data) showed that the major fraction of the turnip cytochrome *f* isolated from the membrane had lost the last 33 amino acids and is a 252-residue polypeptide (Fig. 1b).

*Corresponding authors. †Present address: Institute of Biochemistry, University of Wroclaw, Wroclaw, Poland.

 ISSN 0969-2126

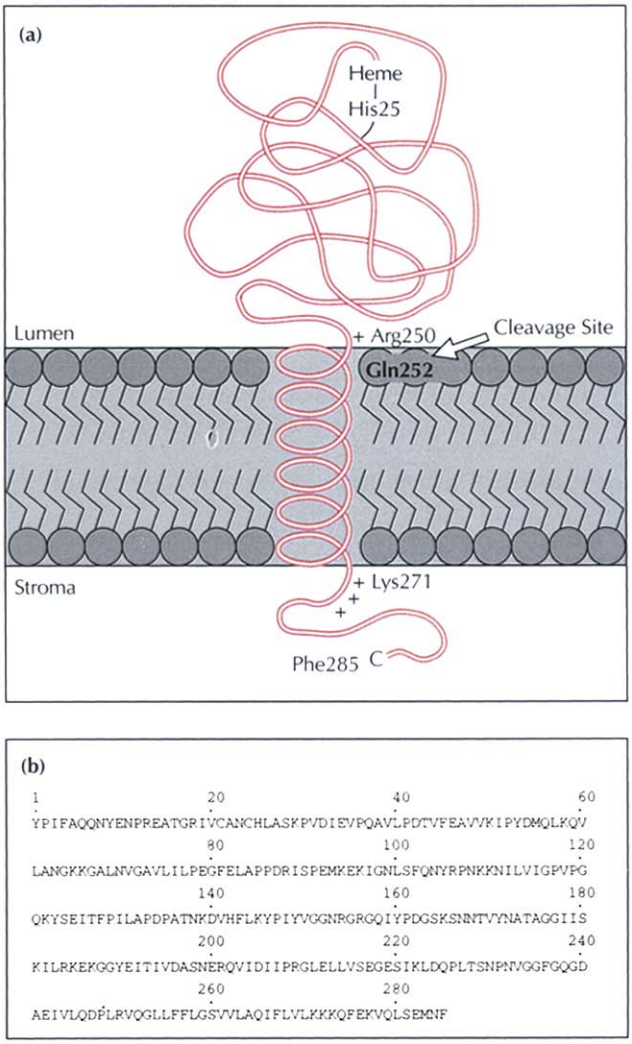

Fig. 1. (a) Schematic membrane topography of cytochrome *f*. The molecule crosses the membrane once, defining three segments of the structure with respect to the chloroplast thylakoid membrane. The amino-terminal 250 residues are on the lumen side of the membrane. The membrane-spanning segment contains the next 20 residues (251–270), which form the hydrophobic anchor. The anchor is punctuated by a tripeptide of conserved lysine residues (271–273) on the stroma side of the membrane and by the conserved Arg250 on the lumen side. The remaining residues (271–285) form a carboxy-terminal carboxypeptidase-accessible [6] segment on the stroma side of the membrane. **(b)** Amino acid sequence of turnip cytochrome *f* [3].

The soluble fragment of turnip cytochrome *f* has the same visible absorbance spectrum (redox-difference peaks at 421, 524, and 554 nm) and midpoint redox potential (0.35–0.38 V) as the intact cytochrome *in vitro* and *in situ* [9]. Its rate of oxidation by the physiological electron acceptor plastocyanin is $2.6 \times 10^7\,M^{-1}\,s^{-1}$ at an ionic strength of 0.3 M [10].

No atomic structures have been reported for the b_6f or bc_1 complexes nor for any subunit of either of these complexes. Thus, it has not been possible to delineate at a molecular level the electron-transfer or proton-translocation pathways for either complex. The electron-transfer reaction between cytochrome *f* and plastocyanin has been studied extensively because the atomic structure of plastocyanin is known [11], but a major obstacle to further progress has been the lack of structural information for cytochrome *f*. We report here the three-dimensional structure determined at a resolution of 2.3 Å of the 252-residue peripheral segment of the membrane-embedded cytochrome *f* from turnip chloroplasts. A report on the crystallization of the cytochrome domain has been presented [8].

Results and discussion

Crystal structure of cytochrome *f*

The crystal structure of the soluble domain of cytochrome *f* was determined by multiple isomorphous replacement and has been refined against diffraction data to 2.3 Å spacings (Tables 1 and 2, Fig. 2). The redox-active lumen-side domain of cytochrome *f* is an elongated molecule of dimensions 75 × 35 × 24 Å (Fig. 3). Two folding domains are positioned end-to-end along the long molecular axis. The smaller folding domain is at the top of the molecule, and the attachment to the membrane anchor peptide is at the bottom in the representation of Fig. 3. The heme is bound by the larger folding domain near the interface with the small domain. The large domain is discontinuous, encompassing residues 1–168 and 232–247. The small domain is composed of residues 169–231.

Large domain structure

The large domain of cytochrome *f* is a three-layer structure dominated by β-strands. The heme and heme-binding peptide (residues 1–25) form the front layer of the domain, and the second and third layers are antiparallel β-sheets. The fold is a marked departure from other *c*-type cytochromes, which are predominantly α-helical. There are only three short helices in the cytochrome *f* crystal structure, all in the large domain. The heme-binding peptide comprises two helices and the loop connecting them, and the third helix is in a long loop connecting two β-strands.

Remarkably, the fold of the large domain of turnip cytochrome *f* is like that of the type III domain (FnIII) of the animal protein fibronectin (Fig. 4a). The FnIII fold has been observed in four proteins with no functional or known evolutionary connection to cytochrome *f*. These include two domains of the bacterial chaperone protein PapD [12], the D2 domain of human cell-surface glycoprotein CD4 [13,14], two extracellular domains of the human growth hormone receptor [15], and the third FnIII domain of the human extracellular matrix protein tenascin [16]. The FnIII fold bears an obvious resemblance to the fold of the immunoglobulin constant domain (Ig-C) and is distinguished from the Ig-C fold by the 'sheet switching' of one β-strand. The β-strands of cytochrome *f* have been labeled in accordance with the immunoglobulin and FnIII notation conventions: the back β-sheet contains β-strands A, B, E, and D, and the front sheet is com-

Table 1. Diffraction data and phasing.

Crystal	D_{min} (Å)	#Observations (#crystals)	#Unique reflections	Completeness (%)	R_{sym}(%)[a]	R_{iso}(%)[b]	#Heavy atom sites	Phasing power[c] (acentric/centric)
Native	2.3	52 944 (4)	13 189	94.7	5.1	–	–	–
$UO_2(Ac)_2$	2.8	55 521 (3)	7 810	99.9	5.7	16.4	5	0.7/0.4
$K_3UO_2F_5$	2.8	11 527 (1)	5 795	77.9	5.2	24.2	13	0.8/0.7
$EtHgPO_4$	2.8	28 527 (2)	7 687	97.2	4.6	15.5	4	0.8/0.5
Baker's dimercurial	2.8	18 052 (1)	7 342	93.6	4.6	27.4	5	1.2/0.8

[a] $R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$ where I = observed intensity, $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry-related reflections. [b] $R_{iso} = \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$ where $|F_P|$ = protein structure factor amplitude, $|F_{PH}|$ = heavy-atom derivative structure factor amplitude. [c] Phasing power = rms ($|F_H|/E$), $|F_H|$ = heavy atom structure factor amplitude, and E = residual lack of closure.

Table 2. Model refinement.

Number of non-hydrogen atoms	1968	
Number of solvent sites	113	
Missing residues	251–252	
Discretely disordered residues	none modeled	
Resolution range	6.0–2.3 Å	
	$F > \sigma_F$	$F > 0$
R-factor	19.8 %	20.5 %
Number of reflections	10 903	11 286
Free R-factor	27.7 %	28.4 %
Number of reflections	551	572
Rms deviation in:		
Bond lengths	0.014 Å	
Bond angles	2.99°	
Ramachandran outliers	none	

posed of β-strands C″, C′, C, F, G, and A′. The very short β-strands A′, C″ and D have only three hydrogen bonds apiece. The three hydrogen bonds between strands A′ and G define the only parallel β structure in cytochrome *f*. The seven long β-strands A, B, C, C′, E, F and G comprise the core of the large domain and form the FnIII fold.

The topology of the core of the cytochrome *f* large domain is more similar to the FnIII fold than to either the Ig-C fold or the immunoglobin variable domain (Ig-V) fold (Figs 4a and 4b). The core of long β-strands in cytochrome *f* superimposes well with FnIII folds and with Ig-V and Ig-C domains (Table 3). The three very short β-strands (A′, C″ and D) in the large domain of cytochrome *f* are not found in the FnIII fold. Strands C″ and D of cytochrome *f* are analogous to strands

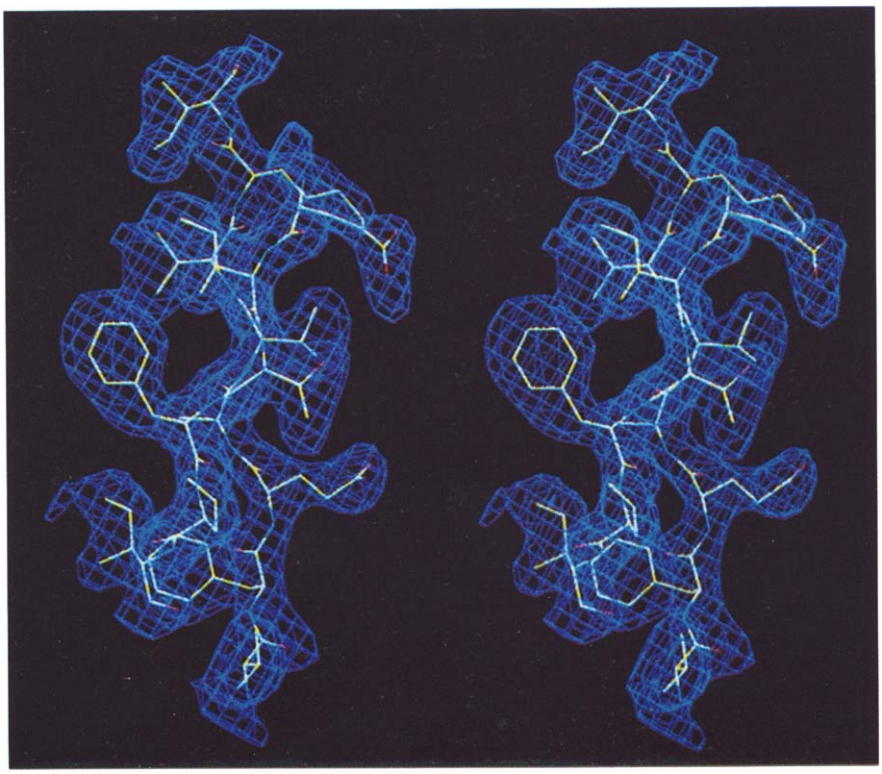

Fig. 2. Stereo diagram of $2F_o$–F_c electron density for β-strands B (residues Val44–Ile51) and E (residues Glu125–Ile130) from the large domain of cytochrome *f*. The 2.3 Å map is contoured at the rms density level. Electron density is continuous for residues 1–248, weak for residues 249–250, and missing for residues 251–252.

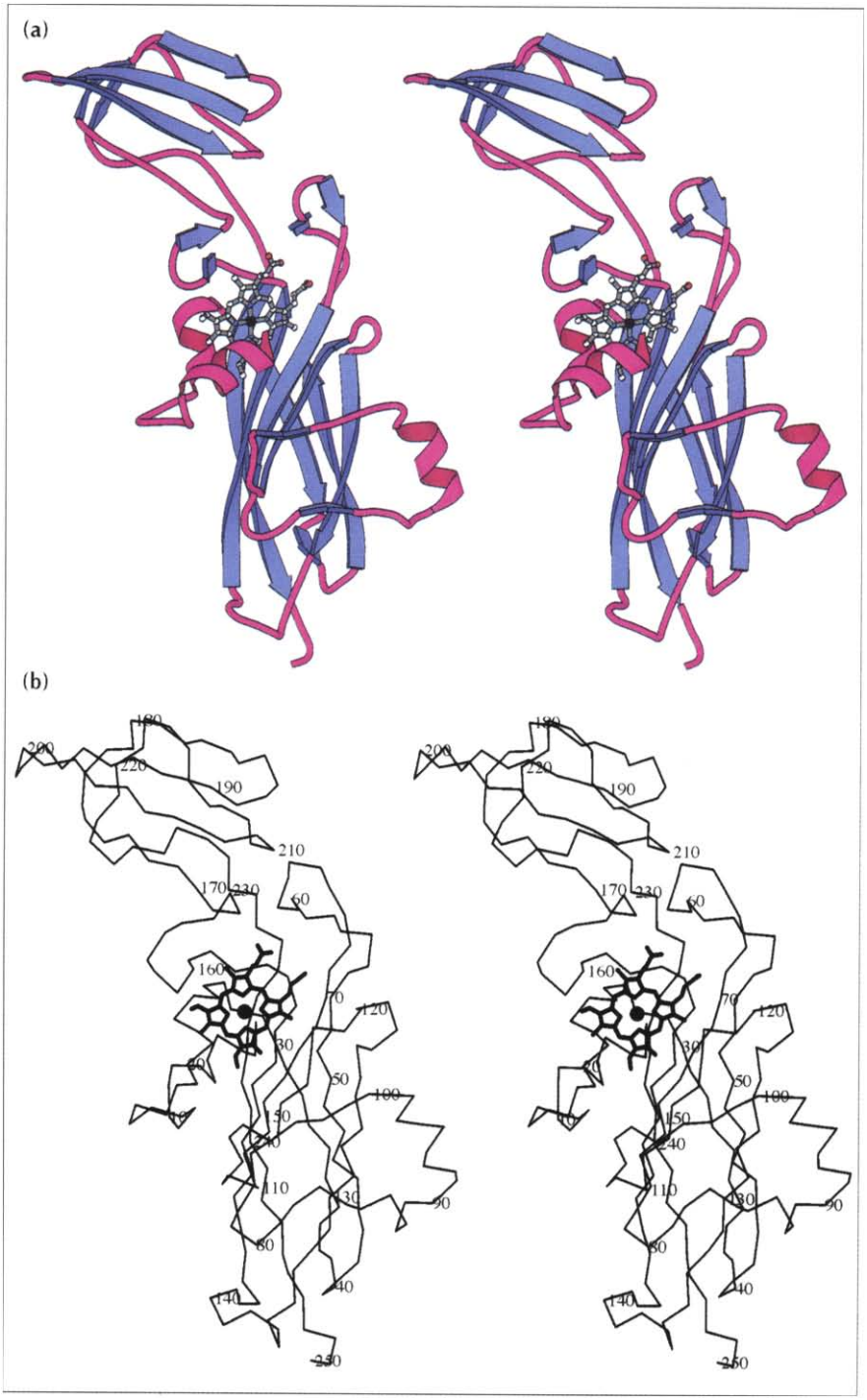

Fig. 3. Polypeptide fold of cytochrome *f*. **(a)** Stereo ribbon diagram with β-strands in blue and α-helices, turns and loops in pink. The heme is shown in 'ball-and-stick' representation. **(b)** Stereo C_α trace with every tenth C_α labeled. (Drawn with Molscript [47].)

C″ and D of the Ig-V domain by virtue of their positions in the β-sheets, but they differ topologically from the Ig-V fold: β-strands occur in the order C D C″ C′ in cytochrome *f* and in the order C C′ C″ D in the Ig-V domain (Fig. 4b). Strands A′, C″ and D of cytochrome *f* are not structurally equivalent to any β-strands of the Ig-V domains of Fab New [17]. The large domain of cytochrome *f* (184 residues) is nearly twice as large as the ~100-residue FnIII fold found in other proteins. The additional mass is due to four peptides: the heme-binding peptide at the amino terminus (25 residues), a long loop between β-strands B and C (20 residues), two β-strands and one helix in the loop connecting strands C and C′ (30 residues) and connections to the cytochrome *f* small domain between strands F and G (15 residues).

Cytochrome *f* sequences are highly conserved across plant and cyanobacterial species [3]. The FnIII sequence motif is widely distributed in animal proteins but has not been detected in any plant protein [18], nor are plant proteins members of the immunoglobulin superfamily [19]. The sequence of the bacterial chaperone protein PapD also does not resemble that

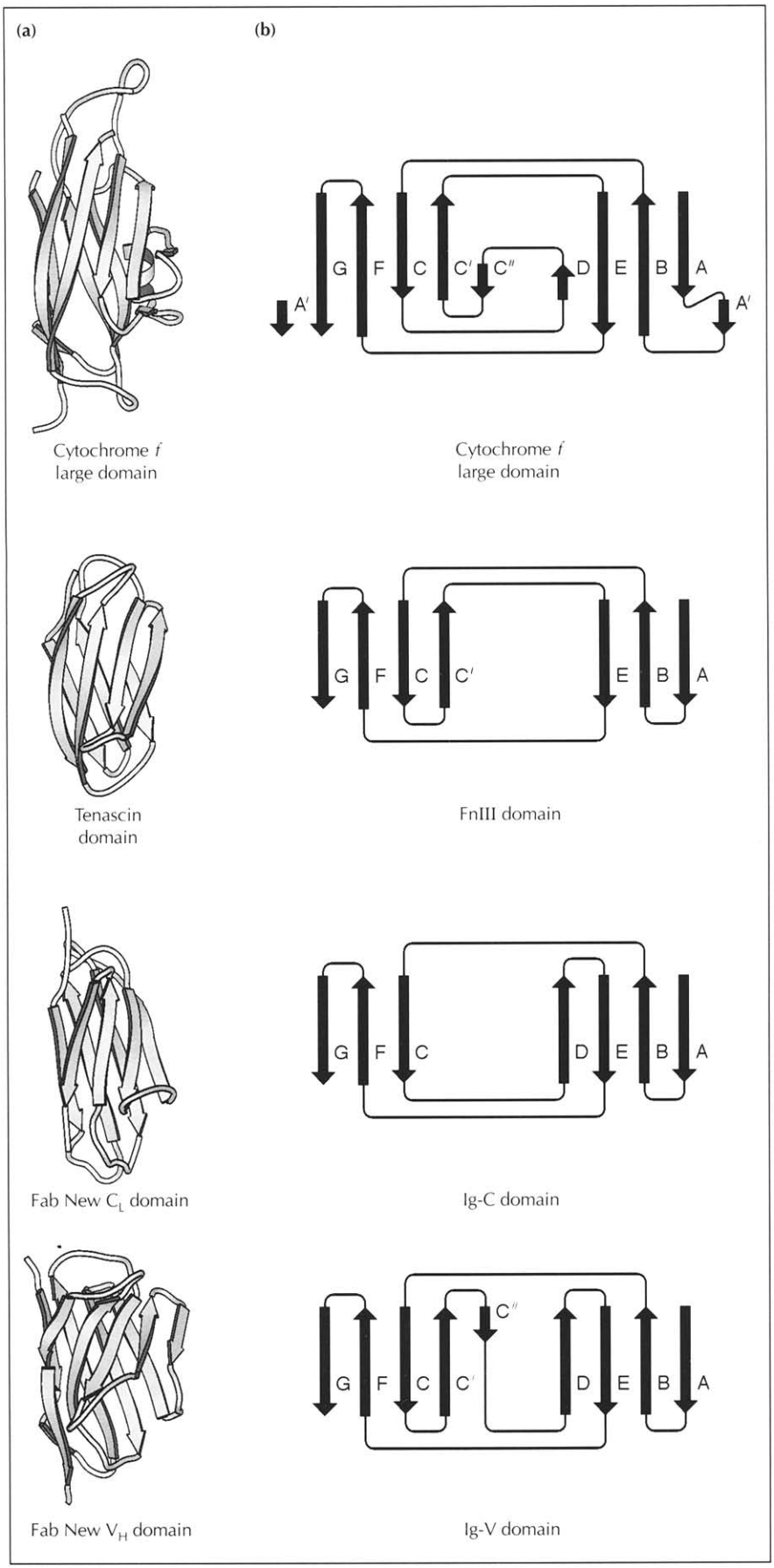

Fig. 4. Comparison of the cytochrome *f* large domain with other domains of similar structure. **(a)** Ribbon diagrams of the core of the large domain of cytochrome *f*, the tenascin third FnIII domain ([16], PDB [48] entry 1TEN), the constant domain of the light chain (C_L) of Fab New ([17], PDB entry 7FAB), and the variable domain of the heavy chain (V_H) of Fab New. (Drawn with Molscript [47].) **(b)** Topology diagrams for each of the domains in (a).

Table 3. Structural similarity of cytochrome *f* large domain with other domains.

Protein	PDB entry	#C_α atoms	Rms deviation[a]
PapD, D1 domain	3DPA	47	1.48 Å
Tenascin, FnIII domain	1TEN	45	1.66 Å
CD4, D2 domain	1CD4	42	1.59 Å
Fab New, V_H domain	7FAB	47	1.54 Å
Fab New, C_L domain	7FAB	41	1.55 Å
Fab New, C_H domain	7FAB	40	1.79 Å
Fab New, V_L domain	7FAB	42	1.79 Å
PapD, D2 domain	3DPA	–	– [b]

[a]Root mean square (rms) deviation between superimposed C_α atoms from the seven core β-strands of the cytochrome *f* large domain and the corresponding C_α atoms of the named domain. [b]The two β-sheets of the D2 domain of PapD could not be superimposed simultaneously with those of cytochrome *f*.

of cytochrome *f*. Therefore, despite the common fold, a common ancestor for cytochrome *f* and any of the other domains in which the FnIII fold has been observed cannot be proposed. It is more reasonable to postulate that the FnIII fold arose in the large domain of cytochrome *f* independently from its origins in the FnIII family, in the immunoglobulin superfamily, or in PapD.

Small domain structure

The small domain of cytochrome *f* is also dominated by β-strands. In this case the domain is folded into a small 'jelly roll', centered at the turn between residues 197 and 200. All hydrogen-bonded β-strands in the small domain are antiparallel. The small domain occurs sequentially between β-strands F and G of the large domain with a connection that appears to be flexible (Fig. 3). However, there is a small core of hydrophobic side chains at the domain junction that may limit flexibility of one domain with respect to the other. The core is composed of the side chains of Val60 and Pro231 and the C_α of Gly64 in the large domain and the side chains of Pro208 and Leu211 and the C_α of Gly210 in the small domain. There is also a hydrogen bond between the side chain of Asn167 in the large domain and the backbone carbonyl of Arg209. All side chains in the hydrophobic interface and Asn167 are conserved in cytochromes *f*.

Unusual heme ligation and heme environment

The active center heme nestles between two short helices at the amino terminus of cytochrome *f*. The second helix contains the sequence fingerprint for the *c*-type cytochromes (Cys-X-Y-Cys-His, residues 21–25). As expected for a cytochrome of the *c*-type, the heme is covalently attached to the protein through thioether bonds to Cys21 and Cys24, and His25 is the fifth heme iron ligand. However, the sixth heme iron ligand is the α-amino group of Tyr1 (Fig. 5). Axial heme iron ligation by the free amino group of the amino-terminal residue of a protein is unprecedented. Amino ligation was expected for cytochrome *f*, but not from the amino terminus of the polypeptide. The absence of an absorbance band at 695 nm for the oxidized protein indicated that methionine was not an iron ligand for cytochrome *f* [20], as it is for cytochrome *c* at neutral pH [21]. The methionine iron ligand of cytochrome *c* can be displaced by an amino ligand, presumably a deprotonated lysine side chain, at alkaline pH. Comparison of visible [22] and near infrared [23,24] magnetic circular dichroism, electron paramagnetic resonance (EPR) [23,24] and resonance Raman [25] spectra of Lys–His cytochrome *c* at alkaline pH with those of cytochrome *f* suggested that the sixth ligand of cytochrome *f* was an amino group. The conserved Lys145 was proposed as a likely candidate for this ligand [25]. In retrospect, the spectral studies were correct in attributing the sixth iron ligand to an amino group, although this group is not derived from lysine. The previously proposed Lys145 ligand is 33 Å from the iron. The α-amino group of Tyr1 with a pK of ~7.2 [26] should be a more stable ligand than the ε-amino group of lysine (pK ~10) in terms of its protonation state at neutral pH.

The immediate environment of the heme is predominantly hydrophobic. The front face of the heme is shielded from solvent by Tyr1, Pro2, Phe4, Pro117 and Pro161. The back face is packed against the side chains of Cys21, Cys24, His25, Asn153, Gly157 and the backbone of residues 154–158. Polar groups in this backbone segment are not pointed towards the heme. The A-B edge of the heme (see Fig. 6) sits directly over β-strand C of the front β-sheet in the large domain of cytochrome *f* with a close contact to the side chain of Gly72; side chains of Val74 and Asn70 contact the porphyrin ring and one propionate substituent, respectively. The B-C edge of the heme contacts Ala5, Tyr9, Val20 and Val74, and the C-D edge contacts the backbone of Tyr160 and Pro161. The propionate carboxyl groups of pyrrole rings A and D are accessible to solvent but are also extensively hydrogen bonded to the side chains of Gln59, Asn70, Arg156, Tyr160, and to two ordered water molecules. The $N_{\delta 1}$ atom of the iron ligand His25 is hydrogen bonded to a well ordered water molecule, which itself is hydrogen bonded to the side chain of Asn153 and to two other ordered waters. These are the only polar groups in the heme-binding pocket.

The heme is more shielded from solvent in cytochrome *f* than in cytochromes c_2 [27] and *c* [28]. Among nonpolar porphyrin substituents in cytochrome *f*, only the vinyl methyl group of pyrrole ring C is solvent exposed. The entire C-D edge of the heme is exposed in cytochromes *c* and c_2, although it is the vinyl methyl group of pyrrole ring C in cytochrome *c* that contacts cytochrome *c* peroxidase in the complex of these redox partners [29]. In contrast to the porphyrin ring, the propionate side groups of pyrrole rings A and D in cytochrome *f* are solvent exposed; in cytochromes c and c_2, they are inaccessible.

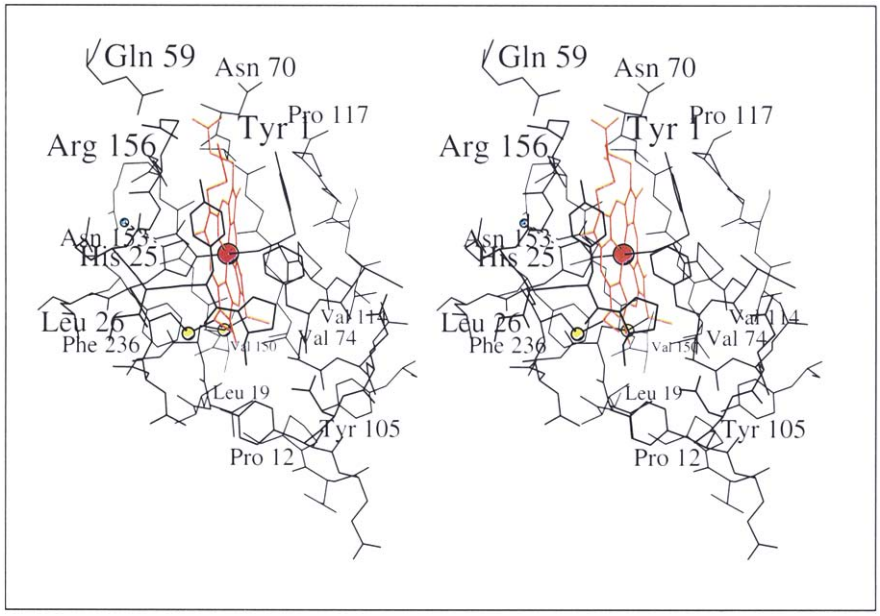

Fig. 5. Stereo diagram showing the environment of the heme (red) in cytochrome *f*. Iron ligands from the protein are the imidazole group of His25 and the α-amino group of Tyr1. The protoporphyrin IX ring is covalently bound through its vinyl groups to Cys21 and Cys24 (sulfurs yellow). The His25 $N_{\varepsilon 2}$–Fe bond distance in the refined model is 2.0 Å. The Tyr1 α-amino N–Fe bond distance is 2.0 Å. The Fe–(Tyr1 N)–(Tyr1 C_α) bond angle is 128°. An ordered water molecule near Asn153 is shown in blue. (Drawn in Molscript [47].)

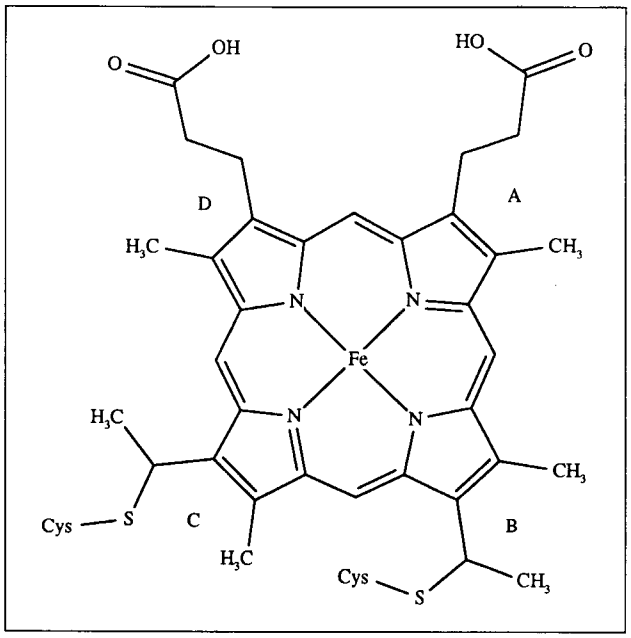

Fig. 6. Schematic diagram of the heme prosthetic group of *c*-type cytochromes. Connections to cysteinyl groups on the protein are shown and pyrrole rings A, B, C and D are labeled.

The heme and the fingerprint peptide are identical in cytochrome *f* and cytochrome *c*, but the heme environments and other segments of protein structure are remarkably divergent (Fig. 7). All common atoms from the fingerprint peptide and heme of cytochrome *f* and *Rhodobacter capsulatus* cytochrome c_2 [27], excluding the propionate carboxyl groups, superimpose with a root mean square (rms) deviation of 0.33 Å. The same atoms of cytochrome *f* superimpose with those of yeast cytochrome *c* [28] with an rms deviation of 0.35 Å. Two classes of *c*-type cytochromes are defined by the mode of heme binding to the fingerprint peptide and differ from one another by a 90° rotation of the histidine ring about the imidazole N–Fe bond and by different side-chain conformations of the cysteine and histidine residues [30]. Cytochrome *f* belongs to the more prevalent class that includes vertebrate, plant, and yeast cytochromes *c* and bacterial cytochromes c_2 and c_{551}. Beyond the ends of the fingerprint peptide, the structures of cytochrome *f* and cytochrome *c* or c_2 are immediately divergent. As expected from the gross differences in the protein structures, there is no conservation of main-chain or side-chain atoms in the heme-binding pocket nor of chemical groups. The only chemical group in the heme environment common to the two proteins, other than the fingerprint peptide, is the phenyl ring of Phe236 of cytochrome *f* (Fig. 5), which is found in an equivalent position and orientation, relative to the heme and the fingerprint peptide, as the side chain of Phe10 in cytochromes *c* and c_2. This residue is conserved in cytochromes *c*, c_2 and *f* [3,21].

Functional significance of the domain structure

The small domain of cytochrome *f* contains a lysine residue, Lys187, which has been cross-linked to Asp44 of plastocyanin, the cytochrome *f* electron acceptor [31]. Lys187 is approximately 28 Å from the heme iron and is located near the top of a long patch of positively charged residues on the surface of turnip cytochrome *f*: Lys187, Arg209, Lys66, Lys65 and Lys58 (Fig. 8). In plastocyanin from the poplar tree, Asp44 is located in a negatively charged surface region about 20 Å from the electron-accepting copper atom [11]. The simplest postulate, and our working hypothesis, regarding the pathway of electron transfer between higher plant cytochrome *f* and plastocyanin is that the small domain of cytochrome *f* interacts with plastocyanin through attraction of oppositely charged patches on the surfaces of the proteins. Comparison of amino acid sequences for cytochrome *f*: plastocyanin partners supports this hypothesis. Except for Arg209, the positively charged

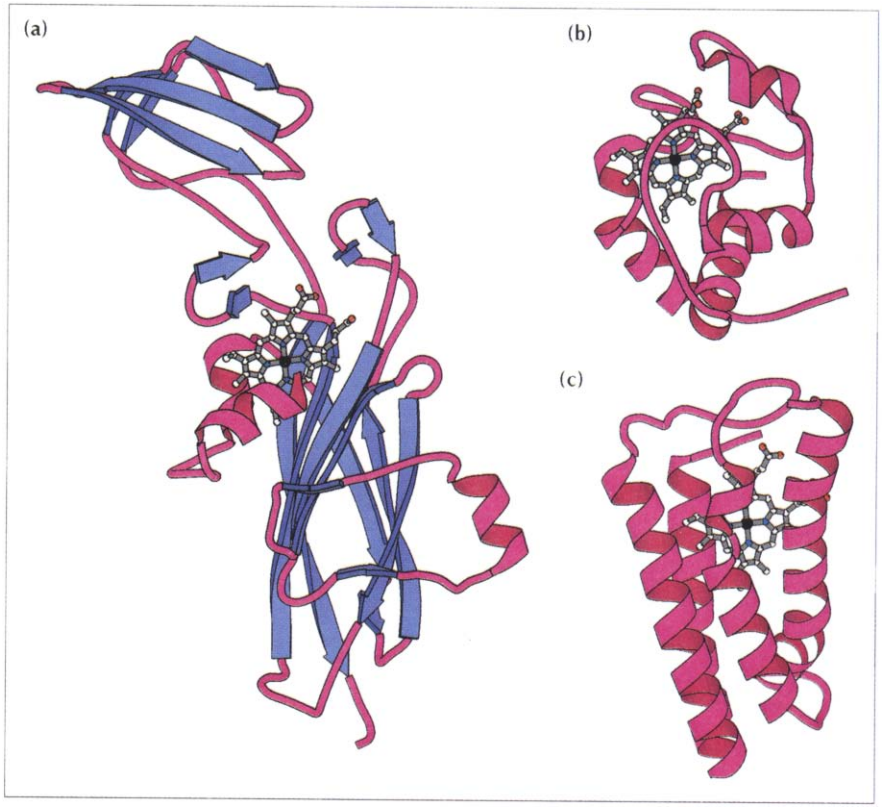

Fig. 7. Comparison of the fold of cytochrome *f* with the known folds of soluble *c*-type cytochromes. **(a)** Cytochrome *f*. **(b)** Yeast cytochrome *c* [28]. **(c)** Cytochrome *c'* [30]. Color scheme as for Fig. 3a. (Drawn with Molscript [47].)

patch on the surface of turnip cytochrome *f* is conserved in the higher plant cytochromes *f*, as is the negatively charged patch on the surface of higher plant plastocyanins. However, the basic patch is totally missing from cyanobacterial cytochromes *f* and partly missing from a green alga sequence [3], and there appear to be positive charges in the area of the acidic patches of some plastocyanins from these organisms [32]. The electrostatic hypothesis is presently of limited use in constructing a model for an electron transport complex between cytochrome *f* and plastocyanin. First, the cross-linked complex may be incompetent in electron transfer [33]. Second, if the connection between domains is at all flexible, then another conformer of cytochrome *f*, different from that in the crystal structure, may exist in the complex with plastocyanin. The large domain of cytochrome *f* contains the beginning of the membrane anchor peptide and therefore must be adjacent to the membrane surface. We hypothesize from the cytochrome *f* structure that the large domain mediates interaction with the Rieske iron–sulfur protein, which is the electron donor associated with the thylakoid membrane. This hypothesis, along with the plastocyanin cross-linking result, implies that the pathway of electron flow is from the redox center of the Rieske protein, through the large domain of cytochrome *f*, to the heme and on to the copper-containing electron acceptor, plastocyanin.

From the protein structure and EPR spectroscopy of oriented membranes containing cytochrome b_6f, we can propose an orientation for the cytochrome *f* molecule with respect to the membrane surface. The cytochrome *f* heme has been found, by EPR spectroscopy, to have a broad distribution of orientations relative to the membrane plane, centered at an angle of 25–30° between the heme and membrane planes [34,35]. The heme plane is tipped approximately 20° from the long axis of the large domain of cytochrome *f*. An oblique and perhaps flexible orientation of the long axis of cytochrome *f* relative to the membrane surface can be inferred. To achieve the heme orientation deduced from the EPR results, the long axis of the large domain of cytochrome *f* must be tipped at least 40° from the perpendicular to the membrane plane. However, flexibility of the tether between the large domain of cytochrome *f* and its membrane anchor may be limited by a very short tether (residues 247–250) that is punctuated on both sides by residues firmly fixed in their respective structures. Immediately preceding the tether, Gln246 terminates β-strand G in the core of the large domain; β-strand G, from residues 244–246, is anchored in the structure by hydrogen bonds to β-strand F on one side and to β-strand A' on the other. Following the tether, Val251 is the first residue of the membrane anchor peptide, which is judged from its 20-residue length to be entirely within the non-polar membrane bilayer.

Consequences for protein translocation and assembly

Cytochrome *f* is synthesized in the chloroplast stroma, and the lumen domain is subsequently transported across the chloroplast thylakoid membrane. An important conclusion from the observation of heme iron ligation by the amino-terminal α-amino group is that

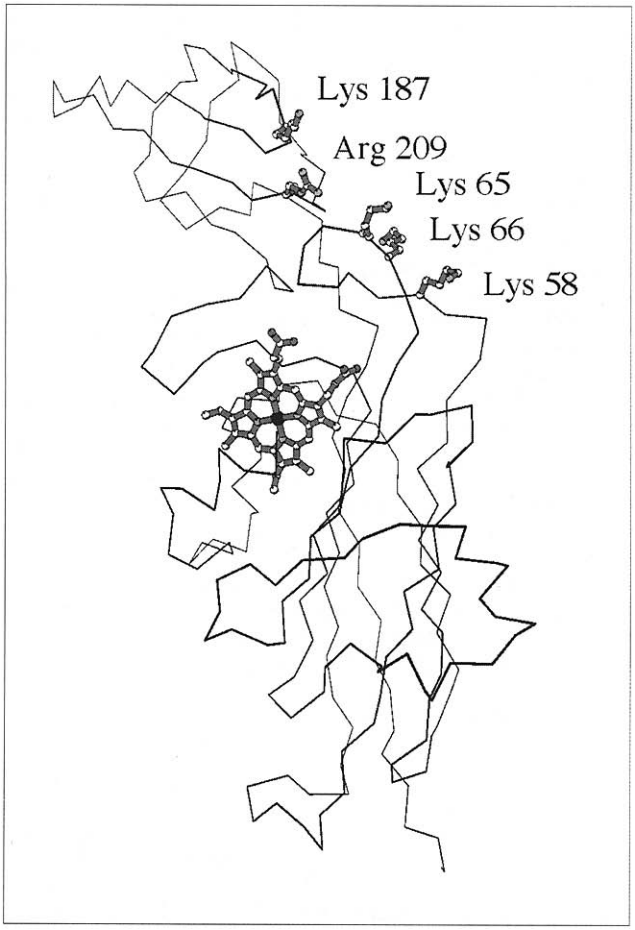

Fig. 8. C_α trace of cytochrome *f* showing the side chains of basic residues near Lys187, which can be chemically cross-linked to Asp44 of plastocyanin [31]. Lys187 and Arg209 form a cluster on one side of the small domain that is adjacent to Lys66, Lys65 and Lys58 on the large domain. All of these basic residues are solvent-exposed. (Drawn with Molscript [47].)

the redox center of cytochrome *f* cannot be assembled until the amino-terminal signal peptide has been removed. This processing event probably occurs after the signal peptide is translocated into the thylakoid lumen because the enzyme activity of the signal peptidase for cytochrome *f* is membrane-associated [36], and because the signal peptidase for the lumenal protein plastocyanin is known to be located on the lumenal side of the membrane [37]. Therefore, assembly of a functional redox center in cytochrome *f* requires that translocation across the membrane proceed sufficiently far to allow cleavage of the cytochrome *f* leader peptide.

The use of the amino-terminal α-amino group as the sixth heme iron ligand can also be viewed as a control mechanism to ensure a proper delay in the onset of protein folding. It is generally accepted that proteins should remain unfolded during their translocation across or into the membrane [38,39]. If heme incorporation, including ligation by the amino terminus of the mature protein, is a critical step in cytochrome *f* folding, then the interdependence of signal peptide cleavage and heme ligation should ensure that cytochrome *f* is incompletely folded before cleavage by the appropriate signal peptidase. This represents a simple biochemical strategy for maintenance of the unfolded state prior to protein translocation.

Relation to cytochrome c_1

The analog of cytochrome *f* in mitochondrial electron transport is cytochrome c_1. Can structural results from cytochrome *f* be applied to cytochrome c_1? The Cys-X-Y-Cys-His heme-binding fingerprint occurs near the amino terminus, and the single membrane anchor segment is near the carboxyl terminus of both proteins. The lumen-side segment of cytochrome *f* is longer than the extra-membrane domain of mitochondrial cytochrome c_1 by approximately 70 residues — the length of the cytochrome *f* small domain. However, the amino acid sequences of cytochromes *f* and c_1 have very low identity apart from the heme-binding fingerprint, as previously noted [40], making sequence alignment unreliable. The situation as regards heme iron ligation is particularly unclear. Residues 1–25 of cytochrome *f* enclose the heme with iron ligation provided by the amino terminus and His25. However, the length and sequence of the peptide preceding the five-residue heme-binding fingerprint is not conserved among cytochromes c_1. Magnetic circular dichroism and EPR spectroscopic studies on the nature of the sixth iron ligand in cytochrome c_1 imply that it is a methionine, as in soluble cytochrome *c*. However, it has been noted that these spectroscopic methods may not be definitive because the set of available reference compounds is limited [23]. In addition, a fusion of *Bradyrhizobium japonicum* cytochromes *b* and c_1 produces a functional cytochrome c_1 [41]. The fusion is between the carboxyl terminus of cytochrome *b* and the amino terminus of cytochrome c_1, suggesting that the amino-terminal residue of cytochrome c_1 may not be a heme iron ligand. Thus, despite segmental and functional similarities between cytochrome c_1 and cytochrome *f*, the extent of structural homology between the two proteins is uncertain.

Biological implications

A great deal of structural information exists for the soluble electron transport proteins cytochrome *c* and plastocyanin, which mediate electron transfer from the cytochrome bc_1 complex to cytochrome oxidase in the electron transport chain of mitochondrial respiration, and from the cytochrome b_6f complex to photosystem I in oxygenic photosynthesis, respectively. This work provides the first atomic structure for any component of either the bc_1 or the b_6f complexes, which are virtually ubiquitous in energy-transducing membranes.

Three aspects of the structure of the 252-residue lumen-side domain of cytochrome *f* are without precedent among the *c*-type cytochromes. The first is the predominance of β structure in the protein fold. The second is the presence of two readily discernible domains. The large, heme-binding domain has a topology identical to that of the type III fibronectin domain, which has not been identified previously in any protein from a plant source. The small domain is distant from the carboxy-terminal membrane anchor and includes a lysine residue that cross-links to plastocyanin, suggesting that it may contribute to the plastocyanin docking site(s). The large domain is immediately adjacent to the membrane anchor and may be associated with binding of the high-potential Rieske iron–sulfur protein, which is the membrane-bound electron donor to cytochrome *f*.

The third unprecedented aspect of the structure is the heme ligation by the α-amino group of the amino-terminal residue of the protein. Ligation of this kind has not previously been described in any cytochrome or other heme-binding protein. Its discovery has implications for the sequence of events associated with import of the cytochrome into the thylakoid membrane, and its folding and assembly. Several translocation events must occur before the folding of functional cytochrome *f* is complete. The signal peptide must first span the thylakoid membrane, exposing its cleavage site on the lumenal side; the soluble segment of the mature protein, including the heme-binding site, must then cross the membrane and the signal peptide must be cleaved. It is generally believed that membrane proteins remain unfolded during translocation. If heme iron ligation is an important step in the folding of cytochrome *f*, then the structure may have revealed a previously undescribed biochemical strategy for maintaining a protein in an unfolded, translocation-competent state.

Materials and methods

The 252-residue lumen-side polypeptide fragment of cytochrome *f*, cleaved after Gln252, was prepared from turnip chloroplasts according to the published protocol [7], except for the addition of protease inhibitors (0.5 mM phenylmethylsulfonyl fluoride, 2 mM benzamidine, 2 mM ε-aminocaproic acid) to the resuspension medium after acetone precipitation. Fractions from the final hydroxyapatite column with OD_{554}:OD_{280} ratios >0.95 were collected, concentrated, and stored at −70°C in 1 mM dithiothreitol, which was present during and after the first chromatography step.

Crystals were grown as described by Martinez *et al.* [8]. The space group is $P2_12_12_1$ with a = 79.2 Å, b = 81.9 Å, c = 46.3 Å, and one cytochrome *f* molecule in the asymmetric unit. The solvent content of the crystals is 54% (by volume) assuming a partial specific volume of 0.75 $cm^3 g^{-1}$ for the protein. Crystals diffract to at least 2.3 Å spacings at room temperature. Mercury derivative crystals were prepared by chemical modification of the protein [42] in the crystal followed by conventional soaking. Uranium derivative crystals were prepared by soaking.

All data were measured with conventional CuK_α X-radiation using an Elliot G20 rotating anode operating at 35 kV, 40 mA with a graphite monochromator. Data were collected on a Siemens area detector diffractometer at room temperature (~22°C) from crystals oriented with a principal axis coincident with the oscillation axis. Data were integrated with XDS [43], and all further processing was done with the CCP4 package [SERC (UK) Collaborative Computer Project 4, Daresbury Laboratory, UK, 1979]. The position of the heme iron was established in a native anomalous difference Patterson map. Major sites for each derivative were located by analysis of isomorphous difference Patterson maps, and minor sites were found in difference Fourier maps. Heavy atom parameters were refined using the program MLPHARE [44]. Anomalous scattering data from all derivatives except uranyl pentafluoride were included in the phasing calculation and were also used to determine the correct enantiomorph. The uranium derivatives had heavy atom sites in common, as did the mercury derivatives. The mean figure of merit for acentric/centric phases to 2.8 Å was 0.69/0.74.

The 2.8 Å multiple isomorphous replacement anomalous scattering (MIRAS) map had a clear definition of solvent and protein. The main chain could be traced without ambiguity. No phase refinement was done. Program O was used for model building [45]. The R-factor for the initial model was 45.3%. This was reduced to 24.2% in one round of simulated annealing with X-PLOR [46]. Further refinement against the 2.3 Å data was also done with X-PLOR. The current model has a real-space correlation coefficient [45] with the MIRAS map of 0.82. This value was 0.78 for the initial model. The model has been deposited with the Brookhaven Protein Data Bank.

Structure-based sequence alignment of cytochrome *f* with other proteins was accomplished by visual superposition of C_α backbone traces (see Table 3). For each pair of structures, superpositions were then determined for all aligned C_α atoms common to the core of seven β-strands. No loop residues were included in the superpositions, and no upper distance cut-off was used to exclude C_α atoms in β-strands.

Acknowledgments. This research was supported by grants from the US Department of Agriculture (91-37306-6317 to JLS), the National Institutes of Health (GM18457 to WAC), and the Lucille P Markey Foundation. The authors thank Dr JC Gray for providing the nucleotide sequence of turnip cytochrome *f* before publication and V Livingston for assistance with the manuscript.

References

1. Cramer, W.A., Furbacher, P.N., Szczepaniak, A. & Tae, G.-S. (1991). Electron transport between photosystem II and photosystem I. In *Current Topics in Bioenergetics.* (Lee, C.P., ed), **16**, pp. 179–222, Academic Press, San Diego.
2. Widger, W.R., Cramer, W.A., Herrmann, R.G. & Trebst, A. (1984). Sequence homology and structural similarity between cytochrome *b* of mitochondrial complex III and the chloroplast b_6f complex: position of the cytochrome *b* hemes in the membrane. *Proc. Natl. Acad. Sci. USA*, **81**, 674–678.
3. Gray, J.C. (1992). Cytochrome *f*: structure, function and biosynthesis. *Photosynth. Res.* **34**, 359–374.
4. Widger, W.R. & Cramer, W.A. (1991). The cytochrome b_6/f complex. In *The Photosynthetic Apparatus: Molecular Biology*

and Operation. (Bogorad, L. & Vasil, I.K., eds), pp. 149–176, Academic Press, San Diego.
5. Horton, P. & Cramer, W.A. (1974). The accessibility of the chloroplast cytochromes *f* and *b*-559 to ferricyanide. *Biochim. Biophys. Acta* **368**, 348–360.
6. Willey, D.L., Auffret, A.D. & Gray, J.C. (1984). Structure and topology of cytochrome *f* in pea chloroplast membranes. *Cell* **36**, 555–562.
7. Gray, J.C. (1978). Purification and properties of monomeric cytochrome *f* from charlock, *Sinapis arvensis* L. *Eur. J. Biochem.* **82**, 133–141.
8. Martinez, S.E., Smith, J.L., Huang, D., Szczepaniak, A. & Cramer, W.A. (1992). Crystallographic studies of the lumen-side domain of turnip cytochrome *f*. In *Research in Photosynthesis.* (Murata, N., ed), **2**, pp. 495–498, Kluwer Academic Publishers, Dordrecht, The Netherlands.
9. Cramer, W.A. & Whitmarsh, J. (1977). Photosynthetic cytochromes. *Annu. Rev. Plant Physiol.* **28**, 133–172.
10. Gross, E.L., Curtiss, A., Durell, S.R. & White, D. (1990). Chemical modification of spinach plastocyanin using 4-chloro-3,5-dinitrobenzoic acid: characterization of four singly-modified forms. *Biochim. Biophys. Acta* **1016**, 107–114.
11. Guss, J.M. & Freeman, H.C. (1983). Structure of oxidized poplar plastocyanin at 1.6 Å resolution. *J. Mol. Biol.* **169**, 521–563.
12. Holmgren, A. & Brändén, C.-I. (1989). Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* **342**, 248–251.
13. Wang, J., *et al.*, & Harrison, S.C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* **348**, 411–418.
14. Ryu, S.-E., *et al.*, & Hendrickson, W.A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* **348**, 419-426.
15. de Vos, A.M., Ultsch, M. & Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312.
16. Leahy, D.J., Hendrickson, W.A., Aukhil, I. & Erickson, H.P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* **258**, 987–991.
17. Saul, F.A., Amzel, L.M. & Poljak, R.J. (1978). Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 Å resolution. *J. Biol. Chem.* **253**, 585–597.
18. Bork, P. & Doolittle, R.F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* **89**, 8990–8994.
19. Williams, A.F. & Barclay, A.N. (1988). The immunoglobulin superfamily — domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381–405.
20. Ho, K.K. & Krogmann, D.W. (1980). Cytochrome *f* from spinach and cyanobacteria: purification and characterization. *J. Biol. Chem.* **255**, 3855–3861.
21. Moore, G.R. & Pettigrew, G.W. (1990). *Cytochromes c: Evolutionary, Structural and Physicochemical Aspects.* Springer-Verlag, Berlin.
22. Siedow, J.N., Vickery, L.E. & Palmer, G. (1980). The nature of the axial ligands of spinach cytochrome *f*. *Arch. Biochem. Biophys.* **203**, 101–107.
23. Simpkin, D., Palmer, G., Devlin, F.J., McKenna, M.C., Jensen, G.M. & Stephens, P.J. (1989). The axial ligands of heme in cytochromes: a near-infrared magnetic circular dichroism study of yeast cytochromes *c*, c_1, and *b* and spinach cytochrome *f*. *Biochemistry* **28**, 8033–8039.
24. Rigby, S.E.J., Moore, G.R., Gray, J.C., Gadsby, P.M.A., George, S.J. & Thomson, A.J. (1988). NMR, EPR and magnetic-CD studies of cytochrome *f*: identity of the haem axial ligands. *Biochem. J.* **256**, 571–577.
25. Davis, D.J., Frame, M.K. & Johnson, D.A. (1988). Resonance Raman spectroscopy indicates a lysine as the sixth iron ligand in cytochrome *f*. *Biochim. Biophys. Acta* **936**, 61–66.
26. Tanford, C. (1962). The interpretation of hydrogen ion titration curves of proteins. *Adv. Protein Chem.* **17**, 69–165.
27. Benning, M.M., *et al.*, & Holden, H.M. (1991). Molecular structure of cytochrome c_2 isolated from *Rhodobacter capsulatus* determined at 2.5 Å resolution. *J. Mol. Biol.* **220**, 673–685.
28. Louie, G.V. & Brayer, G.D. (1990). High-resolution refinement of yeast iso-1-cytochrome *c* and comparisons with other eukaryotic cytochromes *c*. *J. Mol. Biol.* **214**, 527–555.
29. Pelletier, H. & Kraut, J. (1992). Crystal structure of a complex between electron transfer partners, cytochrome *c* peroxidase and cytochrome *c*. *Science* **258**, 1748–1755.
30. Finzel, B.C., Weber, P.C., Hardman, K.D. & Salemme, F.R. (1985). Structure of ferricytochrome *c'* from *Rhodospirillum molischianum* at 1.67 Å resolution. *J. Mol. Biol.* **186**, 627–643.
31. Morand, L.Z., Frame, M.K., Colvert, K.K., Johnson, D.A., Krogmann, D.W. & Davis, D.J. (1989). Plastocyanin cytochrome *f* interaction. *Biochemistry* **28**, 8039–8047.
32. Ho, K.K. & Krogmann, D.W. (1984). Electron donors to P700 in cyanobacteria and algae: an instance of unusual genetic variability. *Biochim. Biophys. Acta* **766**, 310–316.
33. Qin, L. & Kostić, N.M. (1993). Importance of protein rearrangement in the electron-transfer reaction between the physiological partners cytochrome *f* and plastocyanin. *Biochemistry* **32**, 6073–6080.
34. Bergström, J. & Vänngård, T. (1982). EPR signals and orientation of cytochromes in the spinach chloroplast thylakoid membrane. *Biochim. Biophys. Acta* **682**, 452–456.
35. Crowder, M.S., Prince, R.C. & Bearden, A. (1982). Orientation of membrane-bound cytochromes in chloroplasts, detected by low-temperature EPR spectroscopy. *FEBS Lett.* **144**, 204–208.
36. Johnson, E.M., Schnabelrauch, L.S. & Sears, B.B. (1991). A plastome mutation affects processing of both chloroplast and nuclear DNA-encoded plastid proteins. *Mol. Gen. Genet.* **225**, 106–112.
37. Kirwin, P.M., Elderfield, P.D., Williams, R.S. & Robinson, C. (1988). Transport of proteins into chloroplasts: organization, orientation, and lateral distribution of the plastocyanin processing peptidase in the thylakoid network. *J. Biol. Chem.* **263**, 18128–18132.
38. Verner, K. & Schatz, G. (1988). Protein translocation across membranes. *Science* **241**, 1307–1313.
39. Pfanner, N. & Neupert, W. (1990). The mitochondrial protein import apparatus. *Annu. Rev. Biochem.* **59**, 331–353.
40. Hauska, G., Nitschke, W. & Herrmann, R.G. (1988). Amino acid identities in the three redox center-carrying polypeptides of cytochrome bc_1/b_6f complexes. *J. Bioenerg. Biomembr.* **20**, 211–228.
41. Thöny-Meyer, L., James, P. & Hennecke, H. (1991). From one gene to two proteins: the biogenesis of cytochromes *b* and c_1 in *Bradyrhizobium japonicum*. *Proc. Natl. Acad. Sci. USA*, **88**, 5001–5005.
42. Lambert, J.M., Jue, R. & Traut, R.R. (1978). Disulfide cross-linking of *Escherichia coli* ribosomal proteins with 2-iminothiolane (methyl 4-mercaptobutyrimidate): evidence that the cross-linked protein pairs are formed in the intact ribosomal subunit. *Biochemistry* **17**, 5406–5416.
43. Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924.
44. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering.* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–86, SERC, Daresbury, UK.
45. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
46. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460.
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
48. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Received: **6 Dec 1993**; revisions requested: **14 Dec 1993**; revisions received: **20 Dec 1993**. Accepted: **20 Dec 1993**.